

Internet Publishing: Wide Area Information Servers

Brewster Kahle
President
Wide Area Information Servers, Inc
ComNet '94

Overview of Talk

- **Network publishing's brief status**
- **NetBook: An analogy of Internet Information Services**
- **WAIS Inc: a work in progress**
- **Some Upcoming WAIS Systems**

Network Publishing: Where Are We?

- **Networks emerging from Government funding**
 - Networks installed in Academia, Military, and High Tech
 - Networks starting to go commercial
- **Information Services still in Government funding stage:**
 - DARPA/NSF funding still dominant
 - Existing companies adapting (eg. Mead, Dialog)
 - First of Startups (eg. WAIS Inc)
- **Yet to come: Games, Voice/data, Museums, Malls**

What I Really Want

- **My own information to be accessible to me**
- **Published information should find me**
 - Fun to browse
 - Easy to dive into a particular point
- **Usable Anywhere**
- **Help me share what I have learned**

A NetBook Contains:

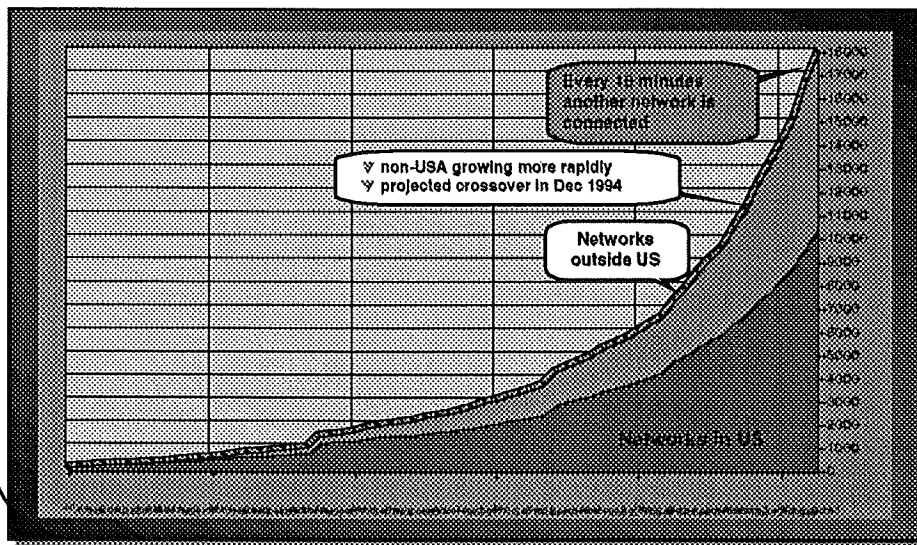
Table of Contents: Browsing (eg. *Gopher*)

Pages: With active footnotes, dynamic illustrations (eg. *WWW*)

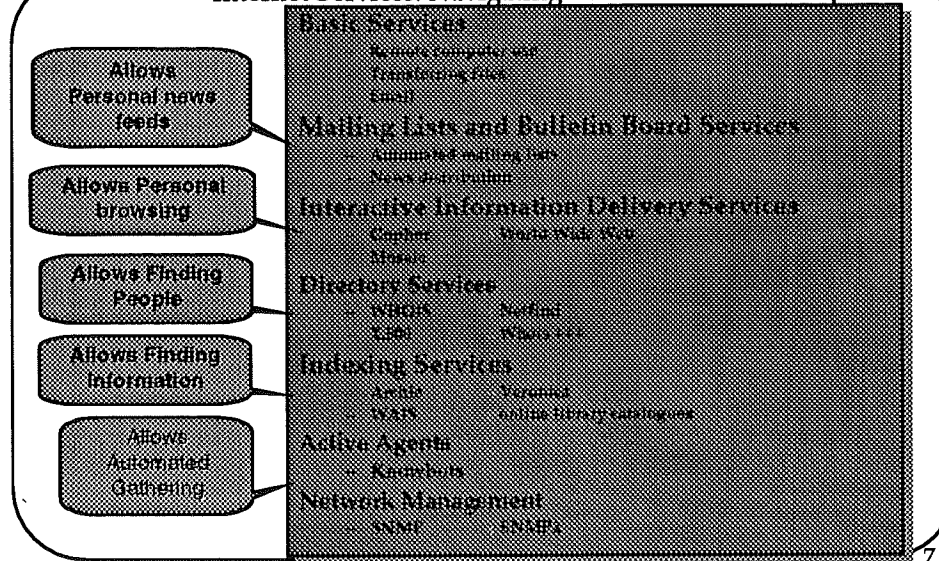
Index: Dive into a subject Area (eg. *WAIS*)

Works over a Network (eg. *Internet*)

History of Connected IP Networks: The Internet



Internet Services: Navigating Global Information Space



7

The Networked Table of Contents

- Browsing
- Hierarchical
- Editor constructed
- New:
 - Cross-machine links to other Tables of Contents
 - Ties to searches, library systems, etc.
 - Problems with reliability

8

Internet Gopher

- Introduced in 1992 by University of Minn
- Thousands of Servers
- Top level gets 10,000 requests an hour
- Very popular with "Internet explorers"

The Networked Page

- Hypertext links: clickable footnote (eg. WWW)
- Dynamic Illustrators: 3-D rotatable, motion, simulation (eg. Interleaf)
- Slower to flip pages: seconds to minutes
- Issues:
 - Fonts & page formats: SGML, Acrobat, RTF, Common Ground, Postscript, ...
 - Picture formats: TIFF, GIF, CCITT(FAX)
 - Different screen sizes, resolutions, network speeds

Flipping Page Speeds

	Size	Over modem	Over Internet	Over Univ Internet
ASCII	2KB	2sec	1sec	1sec
Formatted	30-80KB	30-80sec	5-13sec	1-2sec
w/Picture	150KB	3min	25sec	5sec

Networked Back of the Book Index: Searching

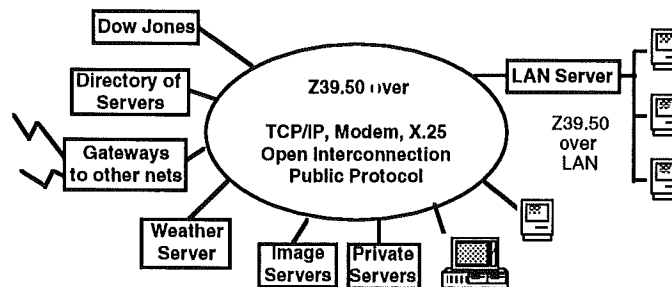
- When you know what you are looking for
- Manual indexing ———> full indexing
- Advanced navigation:
 - topic clustering,
 - query analysis,
 - relevance feedback
- Interactive through gigabytes

WAIS on Internet



- Over 50,000 users on the Internet
- 500 Databases served from 12 Countries: Norway, Canada, UK, ... New database every day.
- Users in 28 Countries: Mexico, Singapore, Finland, Australia, ...
- VT-100 Interface in California: 240 each day, 4,600 different machines

Wide Area Information Server Architecture



- Point-and-click interface to Worldwide Information
- Client -----> Protocol -----> Server

Speed of Searching

	Size	Speed
1 Book	1MB	<1 sec
1 yr of Wall St Jnl	100MB	2 sec
1 Enc Brit	1GB	2-5 sec
1 Lib of Congress	20TB	? sec

Wide Area Information Servers, Inc.

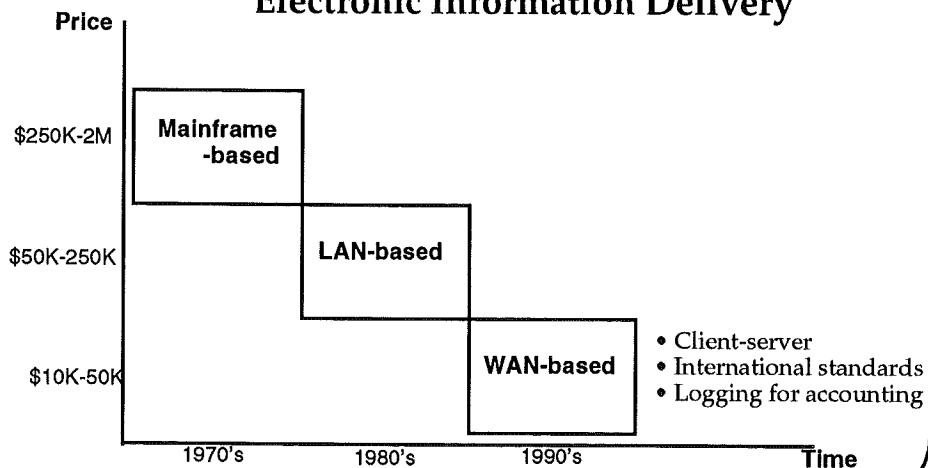
Internet Publishing

Current status

WAIS Inc: Network Publishing

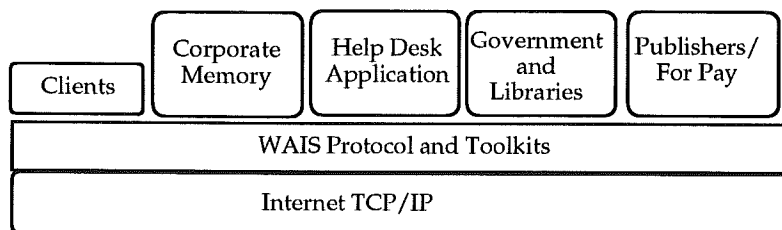
- Founded June 1992 from 1990 Consortium of Apple / Thinking Machines / DowJones / KPMG
- Technology: 2 Founders from Supercomputing
- Internet: Experience in Building an Internet System
- Publishing: 2 Founders from Publishing

Electronic Information Delivery



WAIS Inc. Position

- **WAIS Inc. Brings:**
 - Application Protocols
 - Database Server Technology
 - Wide Area Network Expertise
- **Work with Industry Partners**



WAIS Inc Markets and Partners

- **Large Corporations**
 - Fujitsu, Intel
 - Perot Systems, Mitre
- **Publishers**
 - Dow Jones, encyclopedia, Scholastic Inc
 - Sun Microsystems
- **Government**
 - NASA, Defense, Intelligence
 - US Government Printing Office
- **Libraries**
 - Library of Congress, Rice University

WAIS Inc Products

WAIS Workstation for Unix

\$10k-40k

Novell
Appleshare

WAIS
Workstation

LAN and Internet

WAIS Server for Unix

\$15k-50k

DBMS

WAIS
Server

Internet

Dial-up

Other networks

Services and Training

ComNet '94

21

WAIS Navigation

- Optimized for Wide Area Networks
- Natural Language with Relevance Feedback
- Personal/Enterprise/Global Information
- Interactive Search → Agents

ComNet '94

22

WAIS Inc Server

- **Mixed Natural Language and Boolean**
- **Accounting for Billing**
- **Network Security**
- **Data in original formats**
- **Popular Unix Platforms**

Future Directions

- **Servers: Integrate Gopher, Web, and WAIS Capabilities**
- **Protocols: Working on Standards Committees and Protocol Toolkits**
- **Online Services: Service Bureau of Publishers**

1994 Plans

- **Server 2.0 Release: Integrate Browsing, DBMS, and Security**
- **Protocol Toolkit: Z39.50-v2, URL, (RSA or Kerberos), etc**
- **For-Pay Services on the Internet**

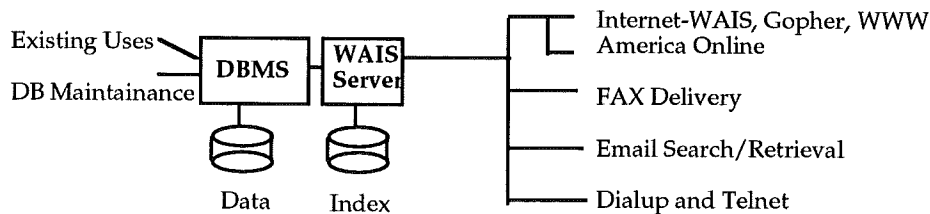
Some WAIS Systems

US Federal Agencies Using WAIS

The OMB Circular A-130 specifically notes: 'Emerging standards such as Wide Area Information Servers (using the NISO Z39.50 standard) will be used increasingly to facilitate dissemination of government data...'

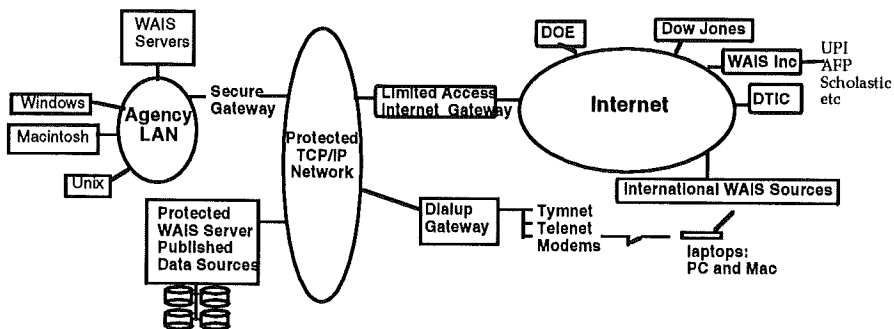
- **NII: Government Information Locators System**
- **Department of Energy**
- **NASA**
- **US Geological Survey**
- **Intelligence (Open Source)**
- **FBI (Terror an Arson)**
- **Defense Technical Information Center (DTIC)**
- **US Department of Agriculture**
- **Environmental Protection Agency**
- **Securities and Exchange Commission (soon)**

Defense Technical Information Center WAIS System

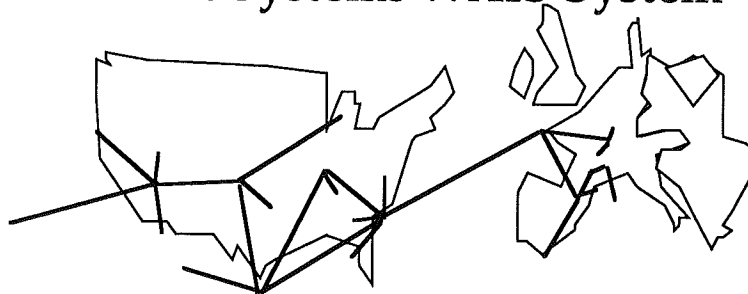


- **Project Just Starting (due within 1 year)**
- **Increased access to Databases and Documents**

US Intelligence Agencies Open Source WAIS System



Perot Systems WAIS System



- **Corporate Information**
 - Resumes, Company Policies, Presentations, Contracts, Proposals, Phone book
- **Published Info**
 - CDROM's, Internet resources

Digital Libraries

- Easy to use card catalog (MARC records)
- Remote use from home or office
- Full text, scanned documents, pictures

```
[pegun.law.columbia.e]  columbia-law-library-catalog  
[pegun.law.columbia.e]  columbia-spanish-law-catalog  
[quake.think.com]       tmc-library
```

Network Publishers: Sun Microsystems

- Combined use of Network Tools to Stay in Touch with Customers
- Email Lists to 10,000 users
- Archive Downloads of 12,000/day
- Press Releases, Bug Fixes, Etc

Digital Libraries and Network Publishing

Large Networks + Too Much Information = **A Problem**

WAIS Can Help Users **Find Critical Information**

WAIS Can Help Publishers **Deliver Information to Users**